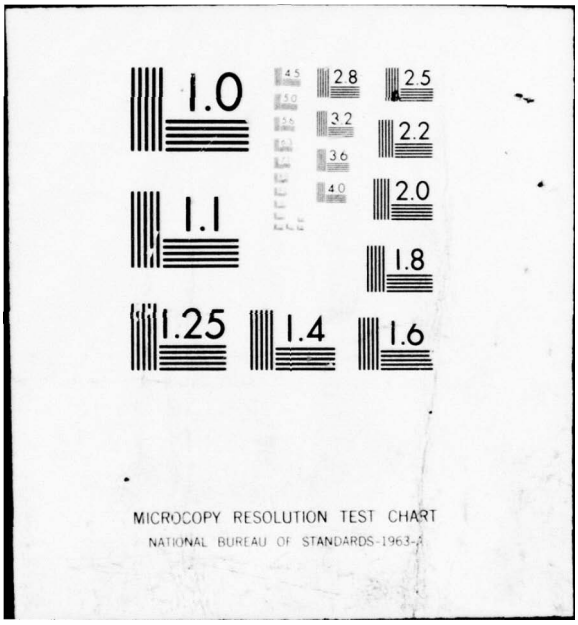
AD-A051 430 STANFORD UNIV CALIF DEPT OF OPERATIONS RESEARCH F/G 12/1
VARIANCE REDUCTION TECHNIQUES FOR THE SIMULATION OF MARKOV PROC--ETC(U)
JAN 78 P HEIDELBERGER N00014-76-C-0578
UNCLASSIFIED TR-44 NL

| OF |
AD
A051 430

END
DATE
FILMED
4-78
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A051430

(14) TR-44

(12)

(6) VARIANCE REDUCTION TECHNIQUES FOR THE SIMULATION OF MARKOV PROCESSES, II. MATRIX ITERATIVE METHODS.

by

(10) Philip Heidelberger*

DDC
RECEIVED
MAR 17 1978
F

(9) TECHNICAL REPORT NO. 44

(11) January 1978

(12) 40 p.

(15) Prepared under Contract N00014-76-C-0578 (NR 042-343)
NSF-MCS-75-23607
for the
Office of Naval Research

Approved for public release: distribution unlimited.

Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government

DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

*The research of this author was also partially supported under National Science Foundation Grant MCS75-23607.

AD No. DDC FILE COPY

402 766

1. Introduction

In this paper we continue the investigation of variance reduction techniques for simulating Markov chains that was begun in Heidelberger (1977). Although a brief review of the notation and results of that paper will be given here, the reader is assumed to be familiar with the contents of the previous paper.

Let $\{X_n, n \geq 0\}$ be an irreducible, aperiodic, positive recurrent Markov chain with finite state space $E = \{0, 1, \ldots, N_E\}$ $(N_E < \infty)$, and transition matrix $P = \{p_{ij}: i,j \in E\}$. It is well known that there exists a probability distribution $\pi = \{\pi_i: i \in E\}$ on E and a random variable X, having distribution π, such that $X_n \Rightarrow X$. π is called the stationary distribution of the Markov chain $\{X_n, n \geq 0\}$. Let f be a real valued function on E and set

$$r = E[f(X)] = \pi f = \sum_{i \in E} \pi_i f(i) \quad . \tag{1.1}$$

We shall be interested in finding r for the given function f. To do so we could solve the system of stationary equations, $\pi = \pi P$, and then find r by applying equation (1.1). However, if the state space is very large it may be quite difficult to solve these equations numerically. In this case it becomes necessary to estimate r via simulation. It is the efficient estimation of such quantities that is our concern. The techniques developed here can also be extended to continuous time Markov chains and

semi-Markov processes by using the techniques of Hordijk, Iglehart and Schassberger (1976). The reader is also referred to equations (3.19) to (3.28) of Heidelberger (1977) for further details concerning this extension.

As in the previous paper we seek to find functions $f_\nu : E \to \mathbb{R}$ so that $r_\nu = \pi f_\nu = r$ for each $\nu = 0, 1, \ldots, k$. By defining

$$\hat{x}_\nu(N) = \frac{1}{N+1} \sum_{n=0}^{N} f_\nu(X_n) , \tag{1.2}$$

it is known that $\hat{x}_\nu(N) \to r_\nu = r$ almost surely (a.s.) as $N \to \infty$ for each ν. Let β be constants such that

$$\sum_{\nu=0}^{k} \beta(\nu) = 1 . \tag{1.3}$$

ACCESSION for
NTIS
DDC
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. SPECIAL
A

If $\hat{x}_\beta(N)$ is defined by

$$\hat{x}_\beta(N) = \sum_{\nu=0}^{k} \beta(\nu)\, \hat{x}_\nu(N) , \tag{1.4}$$

then $\hat{x}_\beta(N) \to r$ a.s. as $N \to \infty$. We now pick $\beta = \beta^*$, where β^* minimizes the asymptotic variance of $\hat{x}_\beta(N)$. We previously studied the variance reductions obtained when the functions f_ν were chosen to be

$$f_\nu = P^\nu f , \qquad \nu = 0, \ldots, k . \tag{1.5}$$

Here alternate methods for generating multiple estimates for $r = \pi f$ are considered in the special case when the Markov chain being simulated

has a finite state space. Once the multiple estimates have been formed, variance reductions can be obtained in exactly the same manner as before. The functions f_ν, $\nu = 0, \ldots, k$ are now found by partially solving an appropriate system of linear equations with some matrix iterative procedure, such as Gauss-Seidel, and then estimating the difference between the true and partial solutions via simulation. The method therefore combines the techniques of numerical analysis and simulation. Each different iterative procedure gives rise to different functions f_ν, and in some cases to different underlying stochastic processes to be simulated.

These methods are quite similar to what are commonly called Monte Carlo techniques for solving systems of linear equations. Monte Carlo solutions for these problems were first suggested by von Neumann and Ulam in the 1940's, however the first published paper on this subject did not appear until 1950 (see Forsythe and Leibler (1950)). There is a vast amount of literature on Monte Carlo methods and the reader is referred to the books by Hammersley and Handscomb (1964), Shreider (1966) or the survey article by Halton (1970) for a complete bibliography.

The motivation and source of problems in the classical Monte Carlo literature are generally quite different from those of stochastic process simulations. The systems of linear equations in the Monte Carlo literature typically arise as finite difference approximations to the solution of multidimensional partial differential equations (this is also the motivation behind much of the work on matrix iterative procedures). As a result many of the matrices involved have special properties, such as being positive definite and symmetric. This type of structure will

generally be absent in the systems of equations which arise in queueing theory or other areas of applied probability. In addition for most Monte Carlo solutions of linear equations, the underlying stochastic process to be simulated arises in a rather arbitrary fashion. Thus simulation often seems to be an unnatural solution technique for these problems. On the other hand, the equations appearing in applied probability have an obvious probabilistic interpretation so that if the standard numerical methods for solving equations are difficult to apply, simulation then becomes a very natural solution technique. It is emphasized that simulation should be used as a last resort; i.e., only after all other methods prove computationally inefficient.

2. Iterative Methods

In this section a class of variance reduction techniques for simulating finite state space Markov chains are described. Pick some state, called the return state, in E, say 0. Set $T_0 = 0$ and for $m \geq 1$ let T_m be the mth time the Markov chain $\{X_n, n \geq 0\}$ visits the return state 0. Call the time between T_{m-1} and $T_m - 1$ the mth cycle and let $\tau_m = T_m - T_{m-1}$ be the length of the mth cycle. Because the Markov chain is assumed to be irreducible and positive recurrent there will be, with probability one, an infinite number of visits to 0 and the expected time between consecutive visits is finite. For any random variable Y let $E_i[Y]$ denote the expectation of Y given that $X_0 = i$. Define

$$y(i) = E_i\left[\sum_{n=0}^{T_1-1} f(X_n)\right] \tag{2.1}$$

and

$$t(i) = E_i[\tau_1] = E_i\left[\sum_{n=0}^{T_1-1} 1\right] . \tag{2.2}$$

Let y and t denote column vectors with ith entries y(i) and t(i) respectively. It is then known (see Crane and Iglehart (1975)) that

$$r = \pi f = y(0)/t(0) . \tag{2.3}$$

Algebraic expressions for y and t have been given in Hordijk, Iglehart, and Schassberger (1976). These expressions will form the basis of the variance reduction techniques.

For any square matrix A let $\rho(A)$ denote the spectral radius of A; i.e., $\rho(A)$ is the modulus of the largest eigenvalue of A. By Theorem 3.7 of Varga (1962) the matrix $I-A$ (where I denotes the identity matrix) is nonsingular if and only if $\rho(A) < 1$. If $\rho(A) < 1$ then

$$(I-A)^{-1} = \sum_{n=0}^{\infty} A^n ,$$

and the infinite series on the right hand side of the equality converges (elementwise). Now introduce the taboo probabilities

$${}_0P_{ij} = \begin{cases} 0 & \text{for } j = 0 \\ p_{ij} & \text{for } j \neq 0 \end{cases} .$$

and let ${}_0P$ be the matrix with entries ${}_0P_{ij}$. P is nonnegative, irreducible and $\sum_{j\in E} p_{ij} = 1$ for all i so that by Lemma 2.5 of Varga (1962), $\rho(P) = 1$. Since $0 \leq {}_0P \leq P$ with ${}_0P_{i0} < p_{i0}$ for some i (because P is irreducible) Lemma 2.3 of Varga (1962) implies that $\rho({}_0P) < 1$. Therefore $(I-{}_0P)$ is nonsingular and

$$(I-{}_0P)^{-1} = \sum_{n=0}^{\infty} {}_0P^n .$$

We are now ready to give expressions for y and t;

$$(2.4) \qquad y = \sum_{n=0}^{\infty} {}_0P^n f = (I-{}_0P)^{-1} f \quad .$$

Furthermore y satisfies the set of linear equations

$$(2.5) \qquad y = f + {}_0Py \quad .$$

If $f = e$, a vector of ones, we obtain expressions for t from (2.4) and (2.5).

Equations of the form (2.5) have a very special structure that lend themselves to at least two different methods of solution. The first is matrix iterative procedures. A comprehensives study of these methods is given in Varga (1962). The second approach is Monte Carlo methods or simulation. In fact it was for equations of exactly this form that von Neumann and Ulam first suggested using simulation.

The approach taken here is a middle ground between these two methods. Suppose we have done k iterations of some matrix iterative procedure in an attempt to solve (2.5). Let y^k be our approximation to y after the kth iteration and let $\epsilon^k = y-y^k$. ϵ^k is then the error in the partial solution y^k. A function g_k can be defined so that

$$(2.6) \qquad \epsilon^k(i) = E_i\left[\sum_{n=0}^{T_1-1} g_k(X_n)\right] \quad .$$

We can then obtain an estimate of the error $\epsilon^k(0)$ by setting $X_0 = 0$, simulating a number of independent cycles, and summing the function g_k over each cycle. An unbiased estimate, $\hat{\epsilon}^k(0)$, for $\epsilon^k(0)$ can be obtained and by setting $\hat{y}^k(0) = y^k(0) + \hat{\epsilon}^k(0)$ we then have an unbiased estimate for $y(0)$ (since $y(0) = y^k(0) + \epsilon^k(0)$). As more iterations are performed; i.e., as k increases, we expect y^k to approach y and g_k to approach 0. Therefore the variance of the estimate for y, $\hat{y}^k(0)$, is also expected to approach 0. This will indeed be the case for any convergent iterative method, however the matter is complicated by the fact that we are really interested in estimates of $r = y(0)/t(0)$, not just estimates of $y(0)$. This issue will be addressed later.

We can improve this procedure (at the cost of additional computer storage) by saving the functions $g_0, \ldots, g_k$, obtaining $k+1$ estimates for r and then taking the minimum variance linear combination as before. More specifically let $g_0, \ldots, g_k$ each satisfy equation (2.6). Define $Y_m(\nu)$ by

$$Y_m(\nu) = y^\nu(0) + \sum_{n=T_{m-1}}^{T_m-1} g_\nu(X_n) , \tag{2.7}$$

then $E_0[Y_m(\nu)] = y(0)$. Since during any cycle there is only one index n for which $X_n = 0$ (that index is T_{m-1} for the mth cycle) we note that

$$Y_m(\nu) = \sum_{n=T_{m-1}}^{T_m-1} f_\nu(X_n)$$

where f_ν is defined by

$$f_\nu(i) = \begin{cases} g_\nu(0) + y^\nu(0) & \text{for } i = 0 \\ g_\nu(i) & \text{for } i \neq 0 \end{cases} . \tag{2.8}$$

From here we proceed exactly as before. Set $Z_m(\nu) = Y_m(\nu) - r\tau_m$, then $E_0[Z_m(\nu)] = 0$. Let

$$\sigma_{ij} = E_0[Z_m(i)\ Z_m(j)] , \qquad 0 \leq i,\ j \leq k \tag{2.9}$$

and let Σ_k be the matrix with entries σ_{ij}. Because E is finite $\sigma_{ij} < \infty$ for all i and j. Σ_k is a symmetric positive semidefinite matrix, which we will assume is positive definite.

Suppose now that $X_0 = 0$ and we simulate the process for M (independent) cycles. Let

$$\hat{r}_\nu(M) = \sum_{m=1}^{M} Y_m(\nu) \Big/ \sum_{m=1}^{M} \tau_m , \qquad \nu = 0, \ldots, k . \tag{2.10}$$

Then $\hat{r}_\nu(M) \to r$ a.s. as $M \to \infty$. Let $\underset{\sim}{B}$ be constants summing to one and define

$$\hat{r}_\beta(M) = \sum_{\nu=0}^{k} \beta(\nu)\ \hat{r}_\nu(M) . \tag{2.11}$$

Then $\hat{r}_\beta(M) \to r$ a.s. as $M \to \infty$ and we can form confidence intervals for r based on the central limit theorem

$$\frac{\sqrt{M}\,(\hat{r}_\beta(M) - r)}{\sigma_k(\beta)/E_0(\tau_1)} \Rightarrow N(0,1) \qquad \text{as } M \to \infty \tag{2.12}$$

where

$$\sigma_k^2(\beta) = \underset{\sim}{\beta}\, \Sigma_k\, \underset{\sim}{\beta}' = \sum_{i=0}^{k} \sum_{j=0}^{k} \beta(i)\, \sigma_{ij}\, \beta(j) .$$

$\underset{\sim}{\beta}$ is now chosen to minimize $\sigma_k^2(\beta)$ and as before

$$\underset{\sim}{\beta}^* = \underset{\sim}{e}\, \Sigma_k^{-1} / \underset{\sim}{e}\, \Sigma_k^{-1}\, \underset{\sim}{e}' \tag{2.13}$$

$$\sigma_k^2(\beta^*) = 1/\underset{\sim}{e}\, \Sigma_k^{-1}\, \underset{\sim}{e}' \tag{2.14}$$

where $\underset{\sim}{e}$ is a vector of ones. Let $R_k^2 = \sigma_k^2(\beta^*)/\sigma_{00}$, a measure of the amount of variance reduction. This technique can also be applied to the point estimate $\hat{x}_\beta(N)$ defined in (1.5).

We now derive expressions for the functions g_ν (and also for f_ν by equation (2.8)) for a number of matrix iterative methods. We have studied the Jacobi, Gauss-Seidel (G-S) and successive overrelaxation (SOR) methods in addition to their block analogs. These methods seem to be the most popular of the iterative procedures when the problem lacks special properties, such as the transition matrix being symmetric or positive definite.

Jacobi's Method

We week the solution, y, to $y = f + {}_0Py$. Because $\rho({}_0P) < 1$ Jacobi's method is known to converge (to y) for any initial y^0 (see Theorem 3.3 of Varga (1962)). Jacobi's iterative procedure is defined by

$$y^k = f + {}_0Py^{k-1} \qquad k \geq 1 \tag{2.15}$$

with y^0 given. This may be written componentwise as

$$y^k(i) = f(i) + \sum_{j=0}^{N_E} {}_0P_{ij}\, y^k(j) \quad .$$

Starting from equation (2.15) we find

$$(I-{}_0P)y^k = f + {}_0Py^{k-1} - {}_0Py^k$$

$$y^k = (I-{}_0P)^{-1}f + (I-{}_0P)^{-1}\,[{}_0P(y^{k-1} - y^k)]$$

$$y^k = y + (I-{}_0P)^{-1}\,[{}_0P(y^{k-1} - y^k)] \ ,$$

the last equality being true by equation (2.4). For $k \geq 1$ define g_k by

$$g_k = -{}_0P(y^{k-1} - y^k) \ , \tag{2.16}$$

then

$$y(i) = y^k(i) + E_i\left[\sum_{n=0}^{T_1-1} g_k(X_n)\right]$$

which is the desired result. If $y^0(i) = 0$ for all $i \in E$, then the formulas simplify to

$$y^k = \sum_{n=0}^{k-1} {}_0P^n f \,, \qquad k \geq 1$$

and

$$g_k = {}_0P^k f \,, \qquad k \geq 0 \,.$$

Gauss-Seidel

The Gauss-Seidel iterative procedure can be considered as an acceleration to Jacobi's method. In Jacobi's method only the values of y^{k-1} are needed to generate y^k, whereas in G-S the most recently available values $y^k(j)$ for $j < i$ are used to find $y^k(i)$. By Theorem 3.3 of Varga (1962) it is known that G-S converges at a faster asymptotic rate than Jacobi's method. The basic G-S iteration is defined by

$$(2.17) \qquad y^k(i) = f(i) + \sum_{j=0}^{i-1} {}_0P_{ij}\, y^k(j) + \sum_{j=i}^{N_E} {}_0P_{ij}\, y^{k-1}(j) \,, \qquad k \geq 1$$

where y^0 is given and sums over empty sets are considered to be 0. From equation (2.17)

$$y^k(i) = f(i) + \sum_{j=0}^{N_E} {}_0P_{ij}\, y^k(j) + \sum_{j=i}^{N_E} {}_0P_{ij}(y^{k-1}(j) - y^k(j)) \,.$$

Define g_k by

$$(2.18) \qquad g_k(i) = - \sum_{j=i}^{N_E} {}_0P_{ij}(y^{k-1}(j) - y^k(j)) , \qquad i \in E .$$

Then in matrix notation

$$y^k = f + {}_0Py^k - g_k$$

$$(I-{}_0P)y^k = f - g_k$$

$$y^k = (I-{}_0P)^{-1}f - (I-{}_0P)^{-1} g_k$$

$$y^k = y - (I-{}_0P)^{-1} g_k \quad .$$

Therefore

$$y(i) = y^k(i) + E_i\left[\sum_{n=0}^{T_1-1} g_k(X_n) \right] .$$

As this is the form needed to apply the variance reduction technique, the appropriate function g_k for G-S is given by equation (2.18).

Block and Point Iterative Methods

We now turn to the study of block and point iterative methods. These methods have the interesting property that the underlying Markov chain to be simulated is generally not the same as the original Markov chain, $\{X_n, n \geq 0\}$. The new Markov chains arise as a special case of the method to be presented in a subsequent paper. Suppose we are again trying

to solve $y = f + {}_0Py$. We now partition the state space into disjoint subsets, or blocks, B_i, $i = 0, \ldots, N_B$; i.e., $E = \bigcup_{i=0}^{N_B} B_i$ and $B_i \cap B_j = \emptyset$, the empty set, for $i \neq j$. For convenience we shall assume that $B_0 = \{0\}$, although this may be relaxed using the techniques of the forthcoming paper. We now partition f and y into $(f_0, f_1, \ldots, f_{N_B})'$ and $(y_0, y_1, \ldots, y_{N_B})'$ where f_i and y_i consist respectively of elements $f(j)$ and $y(j)$ for $j \in B_i$. Similarly ${}_0P$ may be partitioned so that

$$ {}_0P = \begin{pmatrix} {}_0P_{00} & {}_0P_{01} & \cdots & {}_0P_{0N_B} \\ {}_0P_{10} & {}_0P_{11} & \cdots & {}_0P_{1N_B} \\ \vdots & & & \vdots \\ {}_0P_{N_B0} & {}_0P_{N_B1} & \cdots & {}_0P_{N_BN_B} \end{pmatrix} $$

where ${}_0P_{ij} = \{{}_0P_{k\ell} : k \in B_i, \ell \in B_j\}$. Our system of equations then becomes (for the ith block)

$$ y_i = f_i + \sum_{j=0}^{N_B} {}_0P_{ij}\, y_j \quad . \tag{2.19} $$

Letting I_{ii} denote the identity matrix with the number of rows and columns equal to the number of elements in the ith block, equation (2.19) can be simplified;

$$(I_{ii} - {}_0P_{ii})y_i = f_i + \sum_{j \neq i} {}_0P_{ij}\, y_j$$

$$y_i = (I_{ii} - {}_0P_{ii})^{-1} f_i + \sum_{j \neq i} (I_{ii} - {}_0P_{ii})^{-1} {}_0P_{ij}\, y_j \quad . \tag{2.20}$$

Define h_i and ${}_0R_{ij}$ by

$$h_i = (I_{ii} - {}_0P_{ii})^{-1} f_i \tag{2.21}$$

and

$${}_0R_{ij} = \begin{cases} 0 & \text{for } i = j \\ (I_{ii} - {}_0P_{ii})^{-1} {}_0P_{ij} & \text{for } i \neq j \end{cases} \tag{2.22}$$

Equation (2.20) then becomes

$$y_i = h_i + \sum_{j=0}^{N_B} {}_0R_{ij}\, y_j \,, \tag{2.23}$$

or if ${}_0R$ is the matrix with elements ${}_0r_{ij}$ which has been partitioned as

$$
{}_0R = \begin{pmatrix} {}_0R_{00} & {}_0R_{01} & \cdots & {}_0R_{0N_B} \\ {}_0R_{10} & {}_0R_{11} & \cdots & {}_0R_{1N_B} \\ \vdots & & & \vdots \\ {}_0R_{N_B0} & {}_0R_{N_B1} & \cdots & {}_0R_{N_BN_B} \end{pmatrix}
$$

and if $h = (h_0, \ldots, h_{N_B})'$ then (2.23) may be rewritten as

$$
y = h + {}_0Ry \quad . \tag{2.24}
$$

If we now perform Jacobi (G-S) Iterations on equation (2.24) we obtain the block Jacobi (G-S) iterative procedure.

We now turn to the probabilistic interpretation of the matrix ${}_0R$. As was done with ${}_0P$, partition P into the blocks B_i, $i = 0, \ldots, N_B$, and let $P_{ij} = \{p_{k\ell} : k \in B_i, \ell \in B_j\}$. Define

$$
R_{ij} = \begin{cases} P_{ij} & i = 0 \\ 0 & i \neq 0,\ i = j \\ (I_{ii} - P_{ii})^{-1} P_{ij}, & i \neq 0,\ i \neq j \end{cases} \tag{2.25}
$$

($I_{ii} - P_{ii}$ will be nonsingular because P is irreducible) and let $R = \{R_{ij} : 0 \leq i, j \leq N_B\}$. It is then easy to show that the matrix R is a transition matrix; i.e., if R has elements r_{ij} for $i, j \in E$, then $r_{ij} \geq 0$ and $\sum_{j \in E} r_{ij} = 1$ for all $i \in E$. The matrix of taboo

probabilities obtained by setting the 0th column of R equal to 0 and leaving R otherwise unchanged is then equal to ${}_0R$. Define the stopping time S by

$$S = \begin{cases} 1 & \text{if } X_0 = 0 \\ \inf\{n > 0 : X_n \notin B_k\} & \text{if } X_0 \in B_k \text{ and } k \neq 0 . \end{cases}$$

One can then show that

$$r_{ij} = P\{X_S = j | X_0 = i\} \tag{2.26}$$

and that

$$h(i) = E_i\left[\sum_{n=0}^{S-1} f(X_n)\right] . \tag{2.27}$$

From equations (2.24), (2.26) and (2.27) we see that if $\{C_n, n \geq 0\}$ is a Markov chain with transition matrix R and if T'_m is the mth time the process $\{C_n, n \geq 0\}$ enters 0, then

$$y(i) = E\left[\sum_{n=0}^{T'_1-1} h(C_n) | C_0 = i\right] .$$

Similarly if $d(i) = E[S | X_0 = i]$ then

$$t(i) = E\left[\sum_{n=0}^{T'_1-1} d(C_n) | C_0 = i\right] .$$

The numbers $d(i)$ may be obtained from equation (2.21) when $f = e$. Define

$$\delta'_m = \sum_{n=T'_{m-1}}^{T'_m - 1} d(C_n) ,$$

and

$$Y'_m(\nu) = y^{\nu}(0) + \sum_{n=T'_{m-1}}^{T'_m - 1} g_{\nu}(C_n) \quad .$$

Setting $Z'_m(\nu) = Y'_m(\nu) - r\delta'_m$ we have $E_0[Z'_m(\nu)] = y(0) - rt(0) = 0$. Let

$$r'_{\nu}(M) = \sum_{m=1}^{M} Y'_m(\nu) / \sum_{m=1}^{M} \delta'_m \quad .$$

We can now proceed as before (inserting primes into the formulas (2.9) and (2.11) - (2.14) whenever necessary). The important thing to notice is that we are now simulating the Markov chain $\{C_n, n \geq 0\}$ with its transition matrix R rather than the Markov chain $\{X_n, n \geq 0\}$ with its transition matrix P.

Block Jacobi

Let h and ${}_0R$ be defined in (2.21) and (2.22). With y^0 given, the block Jacobi iterations are

$$y^k = h + {}_0Ry^{k-1} , \qquad k \geq 1 \quad .$$

The function g_k, which is derived exactly as in Jacobi's method, is given by

$$g_k = - {}_0R(y^{k-1} - y^k) \quad . \tag{2.28}$$

Block Gauss-Seidel

Let y^0 be given and let h and ${}_0R$ be defined in (2.21) and (2.22). The basic block G-S iteration is

$$y^k(i) = h(i) + \sum_{j=0}^{i-1} {}_0r_{ij}\, y^k(j) + \sum_{j=i}^{N_E} {}_0r_{ij}\, y^{k-1}(j) \ ,$$

and the function g_k is defined by

$$g_k(i) = - \sum_{j=i}^{N_E} {}_0r_{ij}(y^{k-1}(j) - y^k(j)) \ . \tag{2.29}$$

If the blocks are chosen to be singletons; i.e., if $B_i = \{i\}$, then the methods are called point Jacobi and point Gauss-Seidel. The convergence of these block iterative methods can be shown by using Theorem 3.13 of Varga (1962). Again the important thing to emphasize about the block methods is that an entirely different Markov chain must be simulated to apply the technique.

Successive Overrelaxation

Successive overrelaxation may be thought of as an acceleration to Gauss-Seidel. Let ω, the relaxation factor, be given. In order for

SOR to converge it is necessary that $0 < \omega < 2$, (see Theorem 3.5 of Varga (1962)), although this is not a sufficient condition ($\omega = 1$ corresponds to G-S). However, the method does converge for $0 < \omega < \omega_{max}$ where $1 < \omega_{max} \leq 2$. Actually for our purposes it is not really crucial that ω be restricted to this interval since for any value of ω and any finite number of iterations k, the procedure defines valid functions g_k for the simulation. Much work has been done on trying to find the optimum relaxation factor, ω_b; i.e., trying to find that value of ω which maximizes the asymptotic rate of convergence of the iterative procedure. In our case, finding ω_b is not so critical since we are interested in minimizing a variance and not in maximizing the rate of convergence of y^k to y, two entirely different matters. What would be more appropriate here is, for a fixed number of iterations k, to find the value of ω, say ω^*, that minimizes the optimal variance, $\sigma_k^2(\beta^*)$. To do so theoretically would be very difficult indeed so it is recommended that ω^* be estimated from relatively short preliminary simulation runs.

The SOR iterative method requires the diagonal elements of ${}_0P$ to be 0. If $p_{ii} > 0$ for some i, partition E into blocks $B_i = \{i\}$ as in the previous section. As a special case of equation (2.25), we obtain a transition matrix R with entries

$$r_{ij} = \begin{cases} p_{ij} & i = 0 \\ 0 & i \neq 0,\ i = j \\ \dfrac{p_{ij}}{1-p_{ii}} & i \neq 0,\ i \neq j \end{cases}$$

and let ${}_0R$ be defined to be the matrix with entries

$$ {}_0r_{ij} = \begin{cases} 0 & j = 0 \\ r_{ij} & j \neq 0 \, . \end{cases} $$

${}_0R$ has all diagonal elements equal to 0. Let h be defined as

$$ h(i) = \begin{cases} f(i) & i = 0 \\ \dfrac{f(i)}{1-p_{ii}} & i \neq 0 \, , \end{cases} \tag{2.30} $$

which is a special case of equation (2.21). Then $y = h + {}_0Ry$, and we are able to perform SOR iterations on this system of equations. With y^0 given, the basic SOR iteration is

$$ y^k(i) = (1-\omega)\, y^{k-1}(i) + \omega\left[h(i) + \sum_{j=0}^{i-1} {}_0r_{ij}\, y^k(j) + \sum_{j=i+1}^{N_E} {}_0r_{ij}\, y^{k-1}(j)\right]. $$

To obtain $g^k(i)$ write

$$ y^k(i) = (1-\omega)\, y^{k-1}(i) + \omega\left[h(i) + \sum_{j=0}^{N_E} {}_0r_{ij}\, y^k(j) + \sum_{j=i+1}^{N_E} {}_0r_{ij}(y^{k-1}(j)-y^k(j))\right] $$

$$ y^k = (1-\omega)\, y^{k-1} + \omega[h + {}_0R\, y^k + b_k] $$

where

$$(2.31) \qquad b_k(i) = \sum_{j=i+1}^{N_E} {}_0r_{ij}(y^{k-1}(j) - y^k(j)) \quad .$$

Further simplification yields

$$y_k = (I - {}_0R)^{-1}h - (I - {}_0R)^{-1} g_k$$

$$= y - (I - {}_0R)^{-1} g_k$$

with g_k defined by

$$(2.32) \qquad g_k = - [b_k + \frac{(1-\omega)}{\omega} (y^{k-1} - y^k)] \quad .$$

We then have

$$(2.33) \qquad y(i) = y^k(i) + E\left[\sum_{n=0}^{T_1'-1} g_k(C_n) | C_0 = i \right] \quad ,$$

where $\{C_n, n \geq 0\}$ is a Markov chain with transition matrix R. Multiple estimates for r are then produced in the same manner as for the block iterative methods (d(i) is obtained from (2.30) with f(i) = 1). For block SOR formulas (2.32) and (2.33) remain valid provided h and ${}_0R$ are defined as in (2.21) and (2.22). Theorem 3.13 of Varga (1962) implies that block SOR converges for $0 < \omega \leq 1$, and therefore by the continuity of eigenvalues for $0 < \omega < \omega_{max}$ where $1 < \omega_{max} \leq 2$.

3. Numerical Considerations

In this section we discuss some of the problems that might be encountered in the implementation of these methods and present numerical results for a particularly simple Markov chain, the queue length process in a finite capacity M/M/1 queue.

The first problem that one faces with these methods is choosing a return state. As Table 1 indicates, the variance reductions can vary dramatically as the return state changes. This is in contrast to the previous method of picking $f_\nu = P^\nu f$ to generate the multiple estimates for r. For that particular choice of functions, the variance reductions can be shown to be independent of the return state. Preliminary simulation runs could be used to determine a good return state. It seems likely that frequently occuring states (or equivalently states with relatively short expected cycle lengths) will be the best candidates for return state.

The iterative methods also assume that the states are ordered in some manner. Very frequently in simulations each state represents a vector, e.g., a state may represent the queue lengths at various service centers in a network of queues. Since there are many ways to order the states and each different ordering gives rise to different functions $f_0, \ldots, f_k$ (and therefore different variance reductions), the simulator must take care to use an ordering that will give good variance reductions. Again this problem is not encountered when the simulator chooses the functions $f_\nu = P^\nu f$; in that case the order of the states is irrelevant.

Tables 2 through 6 give calculated variance reductions for estimating the expected stationary queue length in the finite capacity M/M/1 queue

for a variety of the methods discussed (this continuous time problem has been transformed into discrete time using the techniques of Hordijk, Iglehart and Schassberger (1976)). The block Gauss-Seidel method (Table 6) gives the best variance reductions for all values of ρ, the traffic intensity. The figures for this method include the variance reduction that is obtained by simulating the Markov chain $\{C_n, n \geq 0\}$ rather than $\{X_n, n \geq 0\}$. That variance reduction is reflected in the R_0^2 values and will be discussed at greater length elsewhere. Due to the work involved in forming the transition matrix for the chain $\{C_n, n \geq 0\}$, block G-S requires more computation to be done before the simulation than the other methods. These extraordinarily good results for block G-S may not be typical because of the very special structure of the M/M/1 queue. Observe that each of the other methods is capable of producing substantial variance reductions, although none dominates the others for all values of ρ. It is perhaps somewhat surprising that for SOR (see Table 5), the value of ω does not have that great an impact on the variance reductions. This is probably because the variance reduction results primarily from the high correlation between the estimates $\hat{r}_0(M), \ldots, \hat{r}_k(M)$ and not from how well the iterative procedure performs; i.e., not from how close y^k is to y. In fact for $\rho = .9$ and $y^0 = 0$ G-S requires over 200 iterations to make $\|y^k - y\| / \|y\| \leq 0.01$ where $\|\cdot\|$ denotes the Euclidean norm. For small values of k, y^k and y tend to differ substantially.

The major difficulty with using these methods will be computing and storing the functions f_ν. Unless the transition matrix is quite sparse the amount of work involved in computing $f_0, \ldots, f_k$ for even

small values of k will probably be too great to justify the use of the methods. The work involved in generating $f_\nu = P^\nu f$ is slightly less than that required to form f_ν for G-S. To perform k G-S iterations requires almost exactly the same amount of work as forming $f, Pf, \ldots, P^k f$. However at the end of the kth G-S iteration one must also calculate $\sum_{j=1}^{N_E} o^{p_{ij}} y^k(j)$ in order to compute $f_k(i)$. For the block methods additional computation must be done to form the necessary transition matrix R and functions h and d. Once these have been formed the work needed for block Jacobi (G-S) is about the same as for Jacobi (G-S).

Since all of the iterative methods we have studied are convergent (for at least some values of ω in the case of SOR), $y^k \to y$ as $k \to \infty$. By examining the functions g_k for each method it is seen that $g_k \to 0$ as $k \to \infty$. Recalling that

$$\sigma_k^2 = \operatorname{var}(Z_m(k)) = \operatorname{var}(Y_m(k) - r\tau_m)$$
$$= \operatorname{var}(y^k(0) + \sum_{n=T_{m-1}}^{T_m - 1} g_k(X_n) - r\tau_m) \quad ,$$

it can be shown that $\sigma_k^2 \to r^2 \operatorname{var}(\tau_m)$, which will in general be positive. In fact for many Markov chains $r^2 \operatorname{var}(\tau_m) > \sigma_0^2$ due to the high positive correlation between $Y_m(0)$ and τ_m. Thus for large values of k, $\hat{r}_k(M)$ tends to be a more variable point estimate for r than $\hat{r}_0(M)$. In fact for the iterative methods we have not been able to show that $\sigma_k^2(\beta^*)$ converges to 0 as $k \to \infty$. This again contrasts to the case when $f_k = P^k f$ where, for a finite state space, σ_k^2 and $\sigma_k^2(\beta^*)$ both converge

to 0. In applications, however, k will usually be small so that for the iterative methods this property should not prove troublesome.

We have so far considered performing iterations on the set of equations $y = f + {}_0Py$, which enables the formation of multiple estimates for the numerator of $r = y(0)/t(0)$. It is also possible to apply the method to the equations $t = e + {}_0Pt$, thereby allowing the formation of multiple estimates for the denominator of r. Letting t^ν denote our approximation to t after ν iterations we could find functions e_ν such that

$$t(i) = t^\nu(i) + E_i\left[\sum_{n=0}^{T_1-1} e_\nu(X_n)\right] .$$

Let

$$\tau_m(\nu) = t^\nu(0) + \sum_{n=T_{m-1}}^{T_m-1} e_\nu(X_n) \quad ,$$

then $E_0(\tau_m(\nu)) = t(0)$. Had we done k_1 and k_2 iterations on y and t respectively we could obtain $(k_1+1)(k_2+1)$ point estimates, $\hat{r}_{ij}(M)$, for r where

$$\hat{r}_{ij}(M) = \sum_{m=1}^{M} Y_m(i) / \sum_{m=1}^{M} \tau_m(j) \quad .$$

Let

$$\sum_{i=0}^{k_1} \sum_{j=0}^{k_2} \beta(i,j) = 1$$

and define

$$\hat{r}_B(M) = \sum_{i=0}^{k_1} \sum_{j=0}^{k_2} \beta(i,j)\, \hat{r}_{ij}(M) \quad .$$

We could then pick the constants $\beta(i,j)$ to minimize the asymptotic variance of $\hat{r}_\beta(M)$. To carry this out in practice we would need to estimate a $(k_1+1)(k_2+1)$ by $(k_1+1)(k_2+1)$ dimensional covariance matrix. Because the estimates $\hat{r}_{ij}(M)$ will usually be quite highly correlated, this covariance matrix is likely to be very ill conditioned, even for moderate values of k_1 and k_2. Thus Σ, $\underset{\sim}{\beta}^*$, and $\sigma^2(\beta^*)$ are likely to be difficult quantities to estimate. Table 7 gives variance reductions for the finite capacity M/M/1 queue when $k_1 = 2$ and $k_2 = 0, 1, 2$ for the Gauss-Seidel iterative procedure. For large values of ρ, the additional variance reductions obtained by performing iterations on t are hardly enough to justify the use of multiple estimates for the denominator, particularly considering the probable difficulty in estimating the covariance matrix. Of course for different functions f it may be more appropriate to use multiple estimates for the denominator rather than the numerator of r (for example in estimating the stationary probability of a particular state).

TABLE 1

Effect of Return State on Variance Reductions for Finite Capacity M/M/1 Queue Using Gauss-Seidel:

$r = E(X)$, $y^0 = 0$

ρ	Return State	R_1^2 / R_1	R_2^2 / R_2	R_3^2 / R_3
.5	0	.0720 .2682	.0408 .2021	.0207 .1440
.5	2	.2222 .4713	.1376 .3709	.0557 .2361
.5	4	.5756 .7587	.5680 .7536	.5674 .7532
.5	6	.5225 .7229	.5034 .7095	.5002 .7073

TABLE 2

Variance Reductions for Finite Capacity M/M/1 Queue Using

$f_\nu = P^\nu f : r = E(X)$

ρ	R_1^2 / R_1	R_2^2 / R_2	R_3^2 / R_3
.25	.1168	.0292	.0073
	.3417	.1709	.0854
.50	.2341	.1121	.0524
	.4838	.3347	.2288
.75	.3328	.1440	.0663
	.5769	.3794	.2574
.90	.6050	.2659	.1148
	.7778	.5156	.3388
.95	.6880	.3425	.1607
	.8294	.5852	.4009
.99	.7404	.4056	.2047
	.8605	.6369	.4525

TABLE 3

Variance Reductions For Finite Capacity M/M/1 Queue Using Jacobi's Method:

$r = E(X)$, $y^0 = 0$, Return State = 0

ρ	R_1^2 / R_1	R_2^2 / R_2	R_3^2 / R_3
.25	.1168	.0291	.0073
	.3417	.1710	.0853
.50	.2341	.1289	.0645
	.4838	.3590	.2540
.75	.3332	.2029	.1574
	.5772	.4504	.3967
.90	.6043	.2772	.1468
	.7774	.5266	.3831
.95	.6878	.3465	.1732
	.8294	.5886	.4161
.99	.7405	.4048	.2087
	.8605	.6362	.4569

TABLE 4

Variance Reductions for Finite Capacity M/M/1 Queue
Using Gauss-Seidel:
$r = E(X)$, $y^0 = 0$, Return State = 0

ρ	R_1^2 / R_1	R_2^2 / R_2	R_3^2 / R_3
.25	.0099	.0015	.0002
	.0997	.0390	.0121
.50	.0720	.0408	.0207
	.2682	.2021	.1440
.75	.2720	.1497	.0813
	.5215	.3869	.2851
.90	.5939	.3881	.1834
	.7706	.6230	.4283
.95	.6789	.4794	.2414
	.8239	.6924	.4914
.99	.7321	.5446	.2963
	.8556	.7380	.5443

TABLE 5

Variance Reductions for Finite Capacity M/M/1 Queue Using SOR: ,

$r = E(X)$, $y^0 = 0$, Return State = 0

ρ	ω	R_1^2 / R_1	R_2^2 / R_2	R_3^2 / R_3
.5	.75	.1217	.0341	.0168
		.3489	.1847	.1296
.5	1.00	.0720	.0408	.0207
		.2682	.2021	.1440
.5	1.25	.0268	.0246	.0212
		.1638	.1567	.1455
.5	1.50	.0151	.0080	.0056
		.1230	.0893	.0747
.5	1.75	.1096	.0261	.0091
		.3310	.1616	.0955
.9	.75	.5951	.3484	.1643
		.7714	.5903	.4054
.9	1.00	.5939	.3881	.1834
		.7707	.6229	.4282
.9	1.25	.5987	.4379	.2155
		.7738	.6618	.4642
.9	1.50	.6174	.5007	.2686
		.7858	.7076	.5183
.9	1.75	.6594	.5786	.3547
		.8120	.7607	.5955

TABLE 6

Variance Reductions for Finite Capacity Queue Using Block G-S:

$r = E(X)$, $y^0 = 0$, Return State = 0, Block Sizes = (1,3,3,3,3,2)

ρ	R_0^2 / R_0	R_1^2 / R_1	R_2^2 / R_2	R_3^2 / R_3
.25	.2942	.0003	1.22×10^{-5}	4.17×10^{-6}
	.5424	.0172	.0035	.0020
.50	.4441	.0075	.0026	1.32×10^{-5}
	.6664	.0864	.0512	.0036
.75	.4037	.0116	.0112	1.36×10^{-4}
	.6353	.1080	.1061	.0117
.90	.3624	.0397	.0133	.0013
	.6020	.1992	.1155	.0360
.95	.3510	.0543	.0127	.0032
	.5925	.2331	.1130	.0567
.99	.3429	.0662	.0127	5.23×10^{-5}
	.5856	.2574	.1126	.0072

TABLE 7

Variance Reductions For Finite Capacity M/M/1 Queue
Using Multiple Estimates for Numerator and Denominator and G-S:
$r = E(X)$, $y^0 = 0$, $t^0 = 0$, Return State = 0

ρ	$\frac{R(2,0)^2}{R(2,0)}$	$\frac{R(2,1)^2}{R(2,1)}$	$\frac{R(2,2)^2}{R(2,2)}$
.25	.0015 .0390	2.80×10^{-5} .0053	2.78×10^{-5} .0052
.50	.0401 .2003	.0092 .0961	.0033 .0579
.75	.1498 .3870	.1238 .3518	.1237 .3518
.99	.5448 .7381	.5366 .7326	.5196 .7208

$R(k_1,k_2)^2$ = Variance reduction when $k_1(k_2)$ G-S iterations are performed on the numerator (denominator)

4. Recommendations

We have presented a class of variance reduction techniques for estimating functionals of the stationary distribution of Markov chains. The methods all require additional computation to be done both before and during the simulation, but if the variance reduction obtained is large enough an overall decrease in computation can be achieved. The methods are likely to be computationally efficient when the transition matrix of the Markov chain is relatively sparse. It is impossible to say a priori which method will yield the best variance reduction for a given Markov chain that is to be simulated. The simulator is advised to experiment with the methods on short preliminary runs to get estimates for the variance reductions. Based on these estimates a particular method (or methods, one need not be limited to only one method) could be chosen. If preliminary experimentation is impossible it is then suggested that the functions $f_\nu = P^\nu f$ be used to generate the multiple estimates. While this may not always produce the best variance reductions possible, it is likely to be the most reliable and easily implemented method. The reliability of this method stems from the fact that the variance reductions are independent of the manner in which the states are ordered and which return state is chosen for the simulation.

Acknowledgement

The author wishes to thank Professors Donald L. Iglehart and Arie Hordijk for their valuable suggestions during this research. The author would also like to thank Gail Lemmond for her expert typing of the manuscript. This research was supported by National Science Foundation Grant MCS75-23607 and Office of Naval Research Contract N00014-76-C-0578.

BIBLIOGRAPHY

[1] CRANE, M.A. and IGLEHART, D.L. (1975). Simulating stable stochastic systems, III: Regenerative processes and discrete-event simulations. Operat. Res. 23, 33-45.

[2] FORSYTHE, G.E. and LEIBLER, R.A. (1950). Matrix Inversion by a Monte Carlo Method. Math. Tables Aids Comput. 4, 127-129.

[3] HALTON, J.H. (1970). A retrospective and prospective survey of the Monte Carlo method. SIAM Rev. 12, 1-60.

[4] HAMMERSLEY, J.M. and HANDSCOMB, D.C. (1964). Monte Carlo Methods. Methuen and Co. Ltd., London.

[5] HEIDELBERGER, P. (1977). Variance reduction techniques for the simulation of Markov processes, I: Multiple Estimates. Technical Report 42, Department of Operations Research, Stanford University.

[6] HORDIJK, A., IGLEHART, D.L. and SCHASSBERGER, R. (1976). Discrete time methods for simulating continuous time Markov chains. Adv. Appl. Prob. 8, 772-788.

[7] SHREIDER, Y.A., ed. (1966). The Monte Carlo Method, The Method of Statistical Trials. Pergamon Press Ltd., Oxford.

[8] VARGA, R.S. (1962). Matrix Iterative Analysis. Prentice-Hall, Inc., Englewood Cliffs, N.J.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 44	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Variance Reduction Techniques for the Simulation of Markov Processes, II: Matrix Iterative Methods		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) PHILIP HEIDELBERGER		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0578
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Operations Research Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR 042-343)
11. CONTROLLING OFFICE NAME AND ADDRESS Statistics and Probability Program Office of Naval Research (Code 436) Arlington, Virginia 20360		12. REPORT DATE January 1978 13. NUMBER OF PAGES 37
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

SEE ATTACHED

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

VARIANCE REDUCTION TECHNIQUES FOR THE SIMULATION OF MARKOV PROCESSES, II: MATRIX ITERATIVE METHODS

by

Philip Heidelberger*

Abstract

Let $\{X_n, n \geq 0\}$ be an irreducible, aperiodic, Markov chain with finite state space E, transition matrix P, and stationary distribution π. Let f be a real valued function on E and define $r = \pi f$. A method of reducing the variance of simulation estimates for r is presented. The method combines the techniques of numerical analysis and simulation by partially solving an appropriate system of linear equations using some matrix iterative procedure and then estimating the difference between the true and partial solutions via simulation. After k iterations of the iterative procedure, functions f_ν, $\nu = 0, \ldots, k$ are defined so that $r = \pi f_\nu$ for each ν. Let $\hat{x}_\nu(N) = \sum_{n=0}^{N} f_\nu(X_n)/(N+1)$ and $\hat{x}_\beta(N) = \sum_{\nu=0}^{k} \beta(\nu)\, \hat{x}_\nu(N)$ where $\sum_{\nu=0}^{k} \beta(\nu) = 1$. Then $\hat{x}_\beta(N) \to r$ a.s. as $N \to \infty$ and β is chosen to minimize the asymptotic variance of $\hat{x}_\beta(N)$. Numerical results for a simple queueing model are presented.


78